4Д

УДК 004.934

Г.В. Дорохина

Институт проблем искусственного интеллекта МОН и НАН Украины, г. Донецк
Украина, sgv@iai.donetsk.ua

Методы пофонемного распознавания, использующие свойства языка и речи

Работа посвящена анализу практической значимости решения проблемы пофонемного распознавания речи; выявлению важных для распознавания свойств языка и речи; разработке на их основе схемы анализа речи и методов пофонемного распознавания речи.

Введение

Сложную не решенную на сегодняшний день проблему распознавания речи считают актуальной с 50-х годов прошлого столетия по ряду причин.

Поскольку параметризованная речь – это многомерный временной ряд, анализируя который принимают решение о произнесённом сообщении, данная проблема относится к области принятия решений на основе анализа временных рядов. Такие задачи распространены в экономике, управлении и т.п.

Пофонемное распознавание выполняется путём анализа множества составных гипотез (последовательностей фонем, последовательностей слов и т.д.), соответствующих произнесению. Новые методы пофонемного распознавания могут повлиять на развитие методов принятия решений на основе анализа составных гипотез.

Кроме того, в речи используется многоуровневая система дублирующих друг друга знаков, что обеспечивает значительный запас помехоустойчивости при передаче сообщений между людьми. Изучение этой многоуровневой системы знаков, способов взаимодействия между знаками различных уровней тесно связано с моделированием интеллектуальной деятельности человека, требует привлечения средств компьютерной лингвистики и других направлений искусственного интеллекта, а успех в этой области будет способствовать развитию средств интеллектуализации компьютерных интерфейсов, ускорению и упрощению процесса человеко-машинного общения.

В последнее время в области речевых технологий можно отметить использование статистических методов описания речи на акустическом, лексическом и более высоких уровнях. Это оправдано, если в основании лежит адекватная модель объекта. Кроме того, неявное использование свойств объекта (что происходит при его статистическом описании) в практических целях не развивает науку о самом объекте. С этой точки зрения *актуальной* является разработка методов пофонемного распознавания, использующих явно заданные свойства языка и речи.

Объект исследования – речь как средство передачи сообщений.

Предмет исследования – модели и методы представления и распознавания речи.

Методы исследования. Методы теории информации и кодирования, положения теорий речеобразования и восприятия речи, акустической фонетики, синтаксиса для анализа речи как объекта.

Целью работы является разработка методов пофонемного распознавания речи, использующих явно заданные свойства языка и речи.

Поставленная цель определила задачи исследования.

1. Выявить важные для распознавания свойства языка и речи.
2. Проанализировать существующие методы распознавания речи, положенные в их основу схемы анализа речи, с точки зрения использования выделенных свойств.
3. Разработать схему анализа речи и методы распознавания речи, использующие явно заданные свойства речи.

Использование свойств языка и речи

Анализ источников позволил выделить следующие свойства языка и речи.

1. Продуктивность (С1) – «возможность продуцировать сколь угодно большое количество информационных сообщений, обладающих разным смыслом» [1].

2. Помехозащищённость (С2) – «сохранение смысловой информации при различных вариантах акустических помех и искажений», «достаточная точность передачи смысловой информации при различных вариантах нарушений (не только патологических, но чаще всего ситуационных) процессов речеобразования и речевосприятия» [1].

3. Параллельное независимое использование сегментного и целостного восприятия слов в потоке речи (С3) – один из механизмов, обеспечивающих помехозащищенность речи [1].

4. Звуки, образованные различными источниками звука (голосовой, шумовой, голосовой и шумовой) [2], хорошо различимы (С4).

5. Звуки речи вариативны и изменчивы. Цельного, однородного звука, соответствующего фонеме, не существует (С5) [3, с. 22].

6. Фонема – минимальная единица языка, которая служит для различения и отождествления значимых единиц языка (С6). Утверждение вытекает из определения фонемы: фонема – группа звуков, обладающих определенной артикуляторно-слуховой общностью, и функционально друг другу не противопоставленных, т.е. не встречающихся в одинаковом положении в фонетической структуре слова [4].

Учёт данных свойств различными методами распознавания речи, а также характеристики этих методов отражены в табл. 1.

Таблица 1 – Характеристики методов распознавания речи

Характеристики / Методы	Возможности метода					Описание уровней языка			СРР	Учёт свойств языка и речи					
	Поэлементное распознавание	Команды малого словаря	Команды среднего словаря	Команды большого словаря	Слитная речь	Акустико-фонетический	Лексический	Семантико-синтаксический		С1	С2	С3	С4	С5	С6
НС	-	+	-	-	-	-	-	-	К	-	±	-	-	-	-
ДВД	-	+	+	-	-	-	-	-	К	-	-	-	-	-	-
СММ	+	+	+	+	+	С	С	С	К	-	-	-	-	-	-
КДП	+	+	+	+	+	С	С	С	Л	-	-	-	-	-	-
АФП	+	+	+	+	+	Я	С	С	Л	-	-	±	+	-	-
ИИ	+	+	+	+	+	Я	Я	Я	-	+	+	+	+	+	+

В табл. 1 использованы следующие обозначения: СРР – системы распознавания речи; СММ – скрытые марковские модели; КДП – метод, сочетающий композицию (К) и динамическое программирование (ДП); НС – нейронные сети; ДВД – метод динамической временной деформации; АФП – методы акустическо-фонетического подхода; ИИ – методы искусственного интеллекта; Я – явное использование свойств языка и речи; С – статистическое описание свойств языка и речи; К – коммерческие системы распознавания речи; Л – лабораторные системы распознавания речи.

Свойства С1 – С6, а также результаты собственных исследований [5], [6] позволили дополнительно сформулировать ряд гипотез относительно свойств речи, необходимых для достижения поставленной цели.

Гипотеза 1: Речь – относительный блочный код (Г1)

а) Относительность кода. В кодировании принимает участие интенсивность и частотные характеристики звука. Кодирование производится не за счёт *предъявления* в частотной области сигналов определённой интенсивности, а за счёт *изменения* спектральных характеристик и интенсивности.

б) Блок речи (БР) – это:

– фрагмент речи, отделённый от остальной речи межфразовыми паузами (дыхательными, интонационными и др.);
– фрагмент речи конечной длины;
– изолированное слово или слитно произносимая фраза.

Гипотеза 2: Синтаксическая структура БР формализуема (Г2)

БР – это изолированное слово или простое распространённое неосложнённое предложение.

Гипотеза 3: Фонема – это единица смыслоразличения, а не единица различения звуков (Г3)

Часть гипотезы Г1, касающаяся относительности кода, предполагает возможность использования относительных порогов (не абсолютных), которые возможно установить, исходя из результатов анализа распознаваемого речевого сигнала (РС). Использование данной гипотезы [7-9] даёт возможность повышения эффективности и устойчивости работы СРР при смене звукозаписывающего оборудования и параметров окружения. Представление речи как последовательности БР, естественным образом отделимых друг от друга во времени, позволяет выявить ограничения на длину и структуру этих единиц, что позволит формализовать их с учётом С4 и Г2 [6]. Третья гипотеза основана на свойствах С5 и С6. Она, в отличие от общепринятой в традиции, предполагает выполнение анализа РС как последовательности фонем (очень размытых в пространстве акустических признаков классов) уже после генерации гипотез о слове или последовательности слов, соответствующих произнесению (смысловых единиц).

На рис. 2 а – в приведены используемые в современных системах распознавания схемы обработки речи.

Хорошо разделимыми в пространстве акустических признаков являются широкие фонетические классы (ШФК) звуков, образованные различными источниками звука (С4): невокализованные смычки ([п, т, к]); невокализованные фрикативные ([ш, щ, ч, с, ц, ф, х]); вокализованные фрикативные ([ж, з, $\overline{\text{дж}}$, $\overline{\text{дз}}$]); остальные вокализованные согласные и гласные. Искажения РС, приводящие к ошибкам в различении ШФК-клас-

сов, делают этот РС нераспознаваемым и для человека. Следовательно, выдвижение или исключение гипотез о произнесении необходимо выполнять по результатам анализа последовательности ШФК звуков РС.

Исходя из сказанного, схема обработки речи должна иметь вид, представленный на рис. 1г. За счёт анализа всех допустимых кандидатов (С2) такая схема обеспечивает параллельное независимое использование сегментного (фонемного) и целостного восприятия потока речи (С3). Ограничение на длину распознаваемого БР позволяет предположить формализуемость фразы как последовательности слов. Анализ речи блоками (Г1) может содействовать переходу в распознавании речи от абсолютных порогов к относительным, вычисляемым по БР, а также к возможности формализации их синтаксической структуры (Г2). Учёт (С1), полученной системой распознавания, зависит от способа организации модуля проверки синтаксической связности последовательности слов. Основанная на С5 и С6, гипотеза Г3 отражает порядок анализа РС – фонемный уровень следует за анализом слов или последовательностями слов.

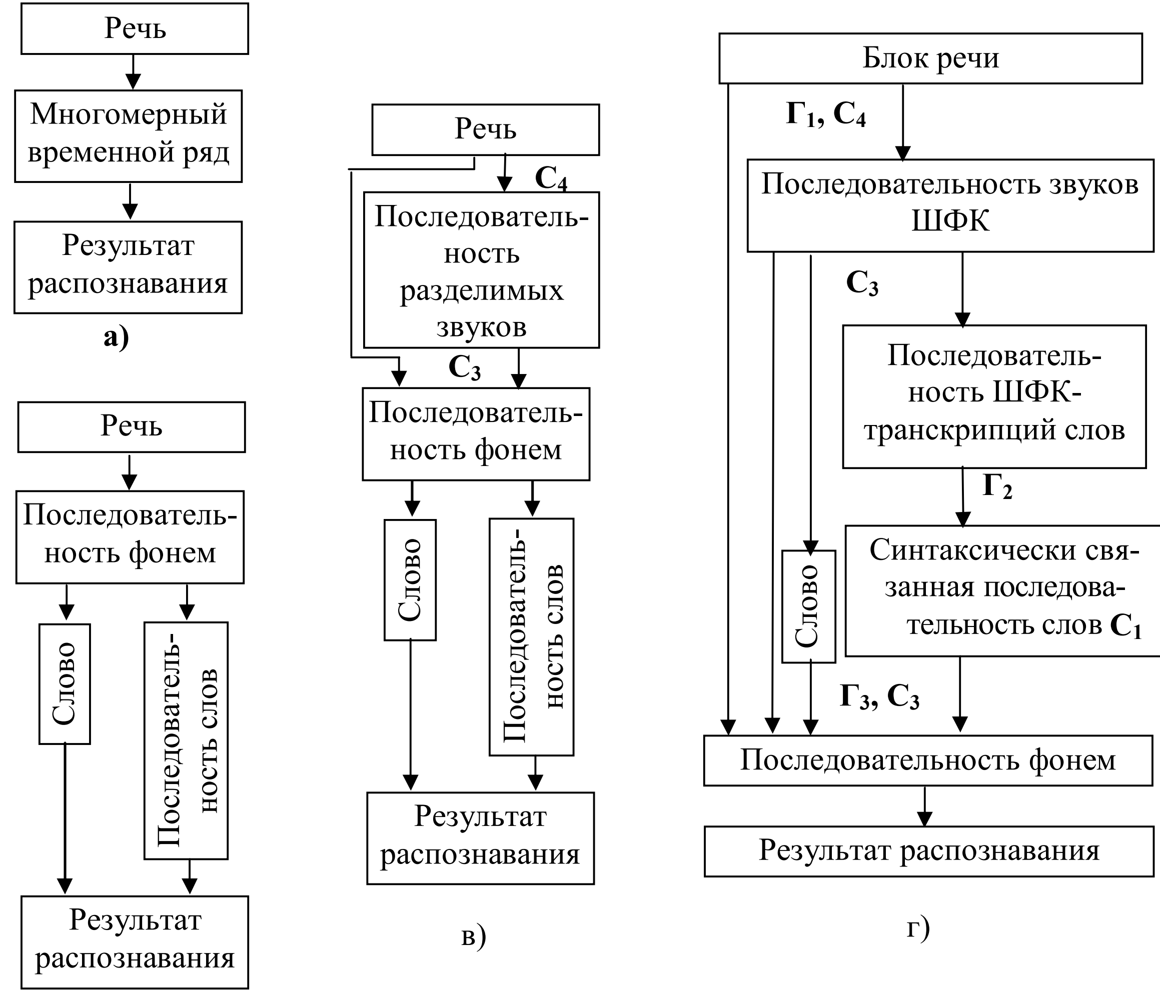


Рисунок 1 – Схемы обработки речи различными методами: а) ДВД, б) СММ, в) методами акустическо-фонетического направления, г) предложенная схема обработки блока речи

Модели пофонемного распознавания речи

Рассматривая слитную речь как последовательность БР, в задаче распознавания получаем две подзадачи: распознавания БР и выбор последовательности БР. Выбор последовательности БР – задача семантического уровня, средствами для решения которой мы не обладаем на данном этапе. Рассмотрим подзадачу распознавания БР.

Ориентируясь на разработку моделей и алгоритмов пофонемного распознавания речи, использующих явно заданные свойства языка и речи, попадаем в область ИИ, где используют методы интеграции источников знаний «снизу-вверх» и «сверху-вниз». С учётом предложенной схемы обработки блока речи разработаны модели пофонемного распознавания «снизу-вверх» (рис. 2а) и «сверху-вниз» (рис. 2б).

На рис. 2 использованы следующие обозначения: БР – блок речи; ШФКТ – транскрипция согласно широкой фонетической классификации; ШФКТС – транскрипция слова согласно широкой фонетической классификации; МИ – морфологическая информация о слове (часть речи, род, число и пр.); ФТ – фонетическая транскрипция; ПО – предметная область; КСГ – контекстно-свободная грамматика, описывающая синтаксис языка.

Данные модели в качестве явно заданных закономерностей речи используют для ограничения количества гипотез фраз ШФК-транскрипции БР. Закономерности языка отражают правила синтаксиса, описывающие синтаксис простых распространённых неосложнённых предложений, а закономерности предметной области – соотношение понятий, выраженное в возможности участия групп слов в определенных видах синтаксической связи.

Распознавание речи при интеграции источников знаний «снизу-вверх» может применяться в задачах диктовки, стенографирования, распознавания спонтанной речи – т.е. в случае, когда трудно заранее оценить перечень возможных фраз и речь нужно фиксировать, а не ставить ей в соответствие некоторое действие. Система, моделирующая этапы «Генерация гипотез последовательностей ШФКТС», «Генерация гипотез последовательностей слов», «Выбор синтаксически допустимых гипотез фразы», описана в работе [5]. При этом множества гипотез (последовательностей ШФКТС, последовательностей слов, фразы) генерируются, анализируются и хранятся в явном виде благодаря способу представления составных гипотез, изложенному в работе [10].

Системы распознавания команд и вопросно-ответные системы предполагают выполнение определённого действия как реакции на результат распознавания. Для обеспечения устойчивой работы системы перечень фраз и действий целесообразно определять до этапа эксплуатации. В таком случае логично применение метода интеграции знаний «сверху-вниз».

По сравнению с распознаванием слитно произносимых фраз согласно методу «снизу-вверх», синтезирующему множество гипотез фразы и анализирующему эти гипотезы, в данном методе выполняется поиск множества гипотез распознаваемой фразы, что требует значительно меньших вычислительных затрат.

Отметим, что для системы распознавания в этом случае не имеет значения распознаёт она слово или фразу – и то, и другое известно системе и представляется последовательностью фонем. То есть для этапа «Пофонемное распознавание» (как для модели «снизу-вверх», так и для модели «сверху-вниз») можно использовать алгоритмы пофонемного распознавания изолированных слов [11]. Фонетическую транскрипцию каждой фразы можем получить с помощью транскриптора, использованного в работе [12], а ШФК-транскрипцию фразы по её написанию – с учётом [5]. Формирование структур словаря фраз, обеспечивающих их скоростной поиск, выполним согласно [10].

Получение ШФК-транскрипции БР можно выполнять, опираясь на работы [7-13]; поиск в словаре фраз множества гипотез фразы, соответствующих ШФК-транскрипции фрагмента речи, обеспечивается сформированными на подготовительном этапе структурами словаря фраз согласно [10].

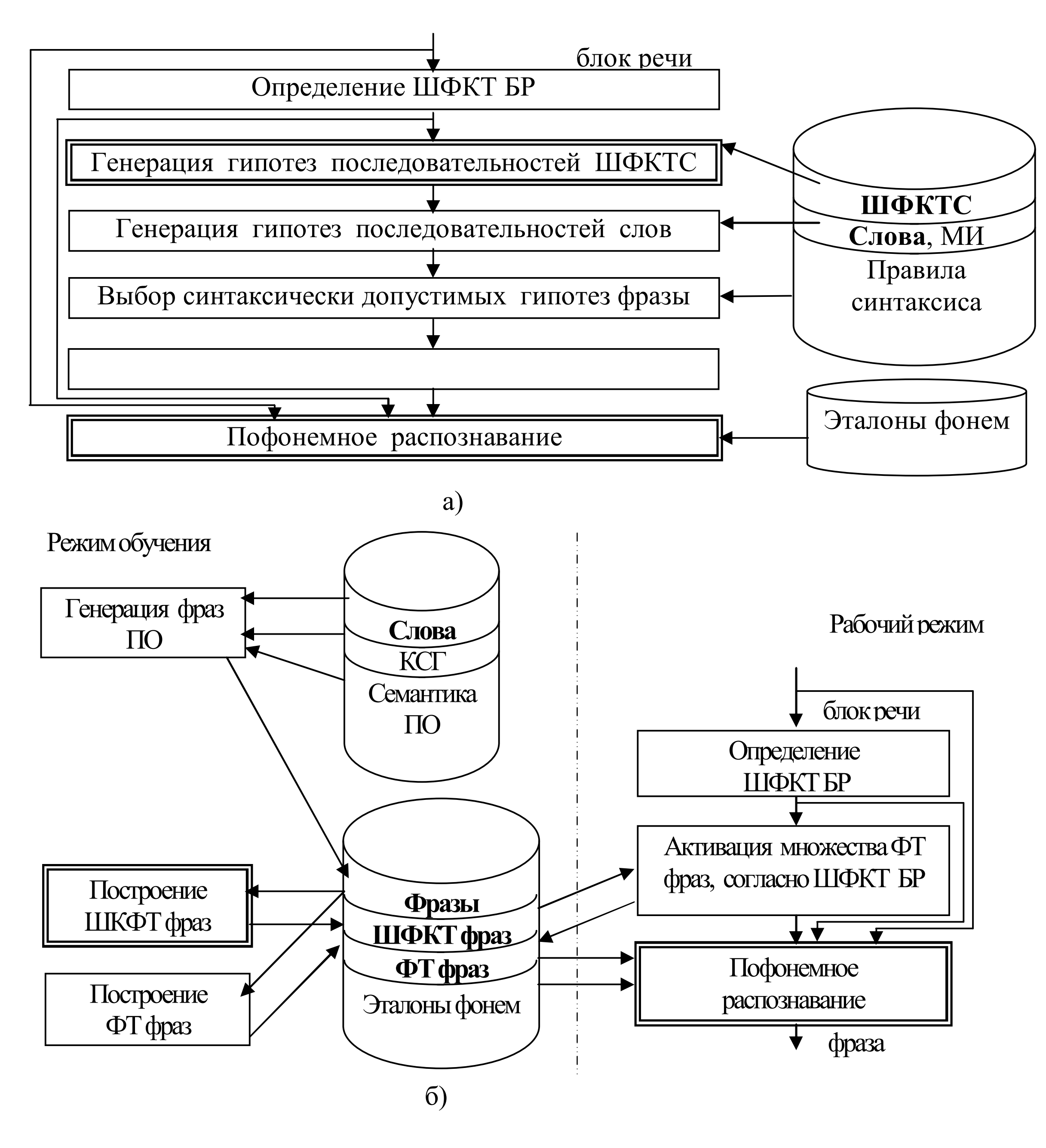


Рисунок 2 – Разработанные модели пофонемного распознавания при методах интеграции источников знаний «снизу-вверх» а) и «сверху-вниз» б)

Следовательно, в указанной модели «снизу-вверх» непроработанным к настоящему моменту остался вопрос синтеза фраз предметной области, чему посвятим дальнейшие работы.

Заключение

В работе проанализирована проблема пофонемного распознавания речи. Выявлены важные для распознавания свойства языка и речи; разработаны на их основе схемы анализа речи и методов пофонемного распознавания речи.

Выявлены важные для распознавания свойства языка и речи, на их основе предложено рассматривать слитную речь как последовательность БР, и сформирована схема обработки блоков речи.

В соответствии с данной схемой разработаны методы пофонемного распознавания, предполагающие интеграцию источников знаний «снизу-вверх» и «сверху-вниз», определены задачи распознавания, в которых они могут быть применены, указаны способы реализации элементов системы пофонемного распознавания.

К отличительным особенностям данных методов относятся явное задание и использование свойств речи и языка; хранение и анализ составных гипотез, представленных согласно [10]; исключение из рассмотрения только недопустимых гипотез, а не анализ наиболее вероятных.

Литература

1. Галунов В.И., Соловьев А.Н. Современные проблемы в области распознавания речи // Информационные технологии и вычислительные системы. – 2004. – № 2.
2. Фант Г. Акустическая теория речеобразования. – М.: Наука, 1964.
3. Ганиев Ж.В. Русский язык: Фонетика и орфоэпия: Учеб. пособие для институтов и фак. иностр. яз. – М.: Высш. шк., 1990. – 174 с.
4. Панов М.В. Современный русский язык. Фонетика: Учебник для ун-тов. – М.: Высш. школа, 1979. – 256 с.
5. Дорохина Г.В. Ограничение количества гипотез фразы при распознавании слитной речи // Известия ТРТУ. – 2005. – № 10. – С. 54-60.
6. Саввина Г.В. Речь как звуковой код // Труды конф. «Информационные технологии в социологии, экономике, образовании и бизнесе» Ялта-Гурзуф (Украина, Крым). – 30 сентября – 9 октября 2003 г. – Запорожье. – 2003. – С. 49-51.
7. Ермоленко Т.В. Разработка системы распознавания изолированных слов русского языка на основе вейвлет-анализа // Искусственный интеллект. – 2005. – № 4. – С. 595-601.
8. Ермоленко Т.В., Фёдоров Е.Е. Методика подавления шума в сигнале, основанная на вейвлет-преобразовании и классификации звуков речи // Сборник научных трудов «Актуальні проблеми автоматизації та інформаційних технологій». – Кафедра вычислительной математики и математической кибернетики ДНУ, 2006. – Т. 10. – С. 50-56.
9. Ермоленко Т.В. Методика формирования эталонов фонем, базирующаяся на вейвлет-преобразовании Морле // Таврический вестник информатики и математики. – 2006. – № 1. – С. 127-132.
10. Пат. України № 78806 «Пристрій для збереження і пошуку рядкових величин та спосіб збереження і пошуку рядкових величин» / Власник: Інститут проблем штучного інтелекту. – Винахідник Дорохіна Г.В. // Промислова власність. Бюл. № 5, 25.04.2007.
11. Козлов А.В., Саввина Г.В., Шелепов В.Ю. Система пофонемного распознавания отдельно произносимых слов // Искусственный интеллект. – 2003. – № 1. – С. 156-165.
12. Грабовая В.А., Федоров Е.Е., Шелепов В.Ю. О системе компьютерного распознавания русской речи с автоматическим построением эталонов // Искусственный интеллект. – 2000. – № 1. – С. 76-81.
13. Ниценко А.В., Шелепов В.Ю. Алгоритмы пофонемного распознавания слов наперед заданного словаря // Искусственный интеллект. – 2004. – № 4. – С. 633-639.

G.V. Dorokhina

Speech Recognition Methods Based on Phoneme Recognition That use Features of Language and Speech

The article is devoted to analysis of speech recognition problem based on phoneme recognition; finding the language and speech features that are useful for speech recognition, designing of speech recognition methods based on phoneme recognition that use features of language and speech.

Статья поступила в редакцию 16.07.2008.